Graphics overlay detection

The invention relates to an overlay detection unit for detecting whether a particular pixel of a first image of a display sequence of video images represents a part of a graphics overlay which is merged with an input sequence of video images to create the display sequence of video images.

The invention further relates to an image processing apparatus comprising:

- receiving means for receiving a signal corresponding to a display sequence of video images;
- an overlay detection unit as described above; and
- an image processing unit for calculating a sequence of output images on basis of the display sequence of video images and on basis of a graphics overlay detection signal being provided by the overlay detection unit.

The invention further relates to a method of detecting whether a particular pixel of a first image of a display sequence of video images represents a part of a graphics overlay which is merged with an input sequence of video images to create the display sequence of video images.

The invention further relates to a computer program product to be loaded by a computer arrangement, comprising instructions to detect whether a particular pixel of a first image of a display sequence of video images represents a part of a graphics overlay which is merged with an input sequence of video images to create the display sequence of video images, the computer arrangement comprising processing means and a memory.

Picture rate conversion, and more particular picture rate upconversion is a well known concept in consumer video equipment like TV sets. See for instance chapter 4 of the book "Video processing for multimedia systems", by G. de Haan, ISBN 90-9014015-8. The purpose of motion compensated picture rate upconversion is to achieve a nice motion portrayal. This is achieved by means of temporal interpolation of a sequence of images on basis of motion vectors being estimated for the images of the sequence. Typically, the different motion vectors are computed for blocks of pixels, e.g. 8*8 pixels. If the images

comprise a graphics overlay, which is typically stationary over time, and also represent moving objects then known picture rate upconversion methods result in break-up of the graphics overlay. See for instance the picture of Fig. 1A.

The graphics overlays might correspond to sub-titles or other graphics information being provided by the content provider or broadcaster. The graphics overlays might also correspond to so-called OSD (On Screen Display) information being generated as part of the graphics user interface of a consumer device like a TV-set, a set top box, a satellite-tuner, a VCR (Video Cassette Recorder) player, a DVD (Digital Versatile Disk) player or recorder. A further option is that the graphics overlays corresponds to meta-data being exchanged in connection with the video data.

The break-up, i.e. distortion of the graphics overlay happens when the graphics overlay covers only a small portion of a block of pixels being used for the motion estimation, while the other portion of that block of pixels corresponds to the moving background. Typically a single motion vector is being estimated for each block of pixels. As a result, this single motion vector does not fit to the actual motion of all pixels in such a block of pixels. If the match error is minimal for the portion of the block of pixels corresponding to the moving background then the estimated motion vector matches to the background motion. Although this causes a relatively large portion of the block of pixels to be correctly motion compensated, this causes distortion of the graphics overlay and hence annoying degradation of the video quality.

It is an object of the invention to provide an overlay detection unit of the kind described in the opening paragraph which is arranged to detect graphics overlays in a robust way.

This object of the invention is achieved in that the overlay detection unit comprises:

- first testing means for testing whether a first difference between a first value of the particular pixel and a second value of a corresponding pixel of a second image of the display sequence of video images is less than a first predetermined threshold;
- second testing means for testing whether a second difference between a third value of a second pixel, being located in a spatial neighborhood of the particular pixel, and a fourth value of a fourth pixel of the second image of the display sequence of video, corresponding to the second pixel, images is less than a second predetermined threshold;

- third testing means for testing whether a third difference between the first value of the particular pixel and the third value of the second pixel is less than a third predetermined threshold; and
- establishing means for establishing that the particular pixel represents the part of the graphics overlay if the first difference is less than the first predetermined threshold, the second difference is less than the second predetermined threshold and the third difference is less than the third predetermined threshold.

In other words, three tests are performed and the three test results are combined to establish whether the particular pixel represents a part of a graphics overlay. A first test is a type of stationary test. That means that it is tested whether the particular pixel and the pixel in a previous or succeeding image, having the same coordinates as the particular pixel, have substantially equal values. The values represent human visible information, e.g. luminance or color. A second test is also a type of stationary test, however not being applied for the particular pixel but for another pixel (called second pixel) in the neighborhood of the particular pixel. With neighborhood is meant a region of e.g. 21*21 pixels. The third test is a type of homogeneity test. That means that it is tested whether the particular pixel and the other pixel (called second pixel) in the neighborhood of the particular pixel have substantially equal values. If all tests are positive, then the particular pixel is assumed to be part of a graphics overlay and is labeled as such. The combination of stationary and homogeneity tests results in a robust detection of graphics overlays.

Preferably the first predetermined threshold and the second predetermined threshold are mutually equal.

An embodiment of the overlay detection unit according to the invention further comprises fourth testing means for testing whether the particular pixel belongs to a group of pixels for which a first motion vector has been estimated which is equal to a null motion vector. By first segmenting the image, for which a graphics overlay detection has to be performed, by means of a motion vector field, a further enhancement of the robustness of the detection is achieved. The segmenting means that blocks of pixels having a motion vector being equal to zero are separated from blocks of pixels having a motion vector which differs from zero. The overlay detection unit is initialized on basis of the motion vector field. The detection of representing a part of a graphics overlay is performed as described above for those pixels of the first image having been assigned a zero motion vector. Afterwards, a kind of growing or dilatation is optionally performed for those pixels having been assigned a non-zero motion vector. The initialization on basis of the motion vector field is advantageous for

the computational complexity and for the robustness of the detection. The motion vector field is preferably computed for the first image but might also have been computed for the second image.

An embodiment of the overlay detection unit according to the invention is arranged to test whether a fourth difference between the first value of the particular pixel and a fourth value of a corresponding pixel of a third image of the display sequence of video images is less than the first predetermined threshold. By means of an additional stationary test on basis of a third image, the robustness of detection is further increased.

An embodiment of the overlay detection unit according to the invention is arranged to test whether a fifth difference between the first value of the particular pixel and a fourth value of a third pixel, being located in the spatial neighborhood of the particular pixel is less than the third predetermined threshold. By means of an additional homogeneity test on basis of a third pixel of the first image, the robustness of detection is further increased. Preferably the particular pixel, the second pixel and the third pixel form a set of mutually connected pixels, or in other words there is a path of adjacent pixels. Testing the condition of connectivity of pixels is advantageous for the robustness.

In an embodiment of the overlay detection unit according to the invention the second pixel is chosen from a set of pixels for which has been established that it represents a further part of the graphics overlay. This embodiment according to the invention is arranged to perform a kind of second iteration over the first image in order to grow, or extend, the set of pixels which have been labeled to represent a part of a graphics overlay, in the first scan over the first image. The quality of detection is further increased by means of the second iteration.

It is a further object of the invention to provide an image processing apparatus of the kind described in the opening paragraph of which the overlay detection unit is arranged to detect graphics overlays in a robust way.

This object of the invention is achieved in that the overlay detection unit comprises:

- first testing means for testing whether a first difference between a first value of the particular pixel and a second value of a corresponding pixel of a second image of the display sequence of video images is less than a first predetermined threshold;
- second testing means for testing whether a second difference between a third value of a second pixel, being located in a spatial neighborhood of the particular pixel, and a

fourth value of a fourth pixel of the second image of the display sequence of video, corresponding to the second pixel, images is less than a second predetermined threshold;

- third testing means for testing whether a third difference between the first value of the particular pixel and the third value of the second pixel is less than a third predetermined threshold; and

- establishing means for establishing that the particular pixel represents the part of the graphics overlay if the first difference is less than the first predetermined threshold, the second difference is less than the second predetermined threshold and the third difference is less than the third predetermined threshold.

The image processing unit might be a temporal up-conversion unit which is arranged to protect the detected parts of the graphics overlay. The image processing unit optionally comprises a display device for displaying the output images.

It is a further object of the invention to provide a method of the kind described in the opening paragraph of which detects graphics overlays in a robust way.

This object of the invention is achieved in that the method comprises:

- testing whether a first difference between a first value of the particular pixel and a second value of a corresponding pixel of a second image of the display sequence of video images is less than a first predetermined threshold;

- testing whether a second difference between a third value of a second pixel, being located in a spatial neighborhood of the particular pixel, and a fourth value of a fourth pixel of the second image of the display sequence of video images, corresponding to the second pixel, is less than a second predetermined threshold;

- testing whether a third difference between the first value of the particular pixel and the third value of the second pixel is less than a third predetermined threshold; and

- establishing that the particular pixel represents the part of the graphics overlay if the first difference is less than the first predetermined threshold, the second difference is less than the second predetermined threshold and the third difference is less than the third predetermined threshold.

It is a further object of the invention to provide a computer program product of the kind described in the opening paragraph of which detects graphics overlays in a robust way.

This object of the invention is achieved in that the computer program product, after being loaded, provides said processing means with the capability to carry out:

- testing whether a first difference between a first value of the particular pixel and a second value of a corresponding pixel of a second image of the display sequence of video images is less than a first predetermined threshold;
- testing whether a second difference between a third value of a second pixel, being located in a spatial neighborhood of the particular pixel, and a fourth value of a fourth pixel of the second image of the display sequence of video images, corresponding to the second pixel, is less than a second predetermined threshold;
- testing whether a third difference between the first value of the particular pixel and the third value of the second pixel is less than a third predetermined threshold; and
- establishing that the particular pixel represents the part of the graphics overlay if the first difference is less than the first predetermined threshold, the second difference is less than the second predetermined threshold and the third difference is less than the third predetermined threshold.

Modifications of the overlay detection unit and variations thereof may correspond to modifications and variations thereof of the image processing apparatus, the method and the computer program product, being described.

These and other aspects of the overlay detection unit, of the image processing apparatus, of the method and of the computer program product, according to the invention will become apparent from and will be elucidated with respect to the implementations and embodiments described hereinafter and with reference to the accompanying drawings, wherein:

Fig. 1A shows an output image of an up-conversion unit according to the prior art;

Fig. 1B shows an output image of an up-conversion unit according to the invention;

Fig. 2A schematically shows two consecutive images of an input sequence of video images;

Fig. 2B schematically shows three consecutive images of an input sequence of video images;

Fig. 3A schematically shows an embodiment of the overlay detection unit according to the invention;

Fig. 3B schematically shows an alternative embodiment of the overlay detection unit according to the invention; and

Fig. 4 schematically shows an embodiment of the image processing apparatus according to the invention.

Same reference numerals are used to denote similar parts throughout the figures.

Fig. 1A shows an output image 100 of an up-conversion unit of the prior art. The output image 100 is based on temporal interpolation of input images. The input images correspond to a part of a movie in which a car is driving in the streets. During the acquisition of the input images the camera was panning in order to keep the car in the center of the consecutive images. As a consequence, the most motion vectors being estimated for the input images are unequal to zero. The output image 100 further comprises graphics overlays which have been created by a DVD player and blended with the images as stored on the DVD. The mixed images, i.e. mix of images as stored and the graphics overlays, are used for the up-conversion. Unfortunately, the up-conversion results in visible artifacts in the graphics overlays 102 and 104. That means that the output image 100 as provided by the up-conversion unit according to the prior art comprises graphics overlays 102 and 104 which are less visible than required. It can be clearly seen that the representation 102 of some textual information, e.g. the string "1:21:24" is distorted. When comparing the output image 100 provided by the up-conversion unit according to the prior art with an other output image 106, as shown in Fig. 1B, being provided by an up-conversion unit according to the invention, it can be observed that the representation 110 of some icons is substantially improved.

Fig. 2A schematically shows two consecutive images 200 and 202 of an input sequence of video images and a number of pixels P(1,n), P(1,n-1), P(2,n) and P(2,n-1) of these images. To determine whether a particular pixel P(1,n) represents a part of a graphics overlay a number of tests are performed, i.e. a number of differences D11, D12 and D22 are computed and compared with respective thresholds. This is described in more detail in connection with Fig. 3A.

Fig. 2B schematically shows three consecutive images 200-204 of an input sequence of video images and a number of pixels P(1,n), P(1,n-1), P(1,n+1), P(2,n), P(2,n-1) and P(3,n) of these images. To determine whether a particular pixel P(1,n) represents a part of a graphics overlay, a number of tests are performed, i.e. a number of differences D11,

D12, D13, D22 and D11p are computed and compared with respective thresholds. This is described in more detail in connection with Fig. 3B.

Fig. 3A schematically shows an embodiment of the overlay detection unit 300 according to the invention. The overlay detection unit 300 is provided with a video signal corresponding to a display sequence of video images. This display sequence of video images might be based on an input sequence of video images which is merged with a graphics overlay. The overlay detection unit 300 is arranged to detect whether a particular pixel of a first image of the display sequence of video images represents a part of such a graphics overlay. The overlay detection unit 300 comprises:

- a first evaluation unit 302 for testing whether a first difference D11 between a first value of the particular pixel P(1,n) and a second value of a corresponding pixel P(1,n-1) of a second image of the display sequence of video images is less than a first predetermined threshold T1;
- a second evaluation unit 304 for testing whether a second difference D22 between a third value of a second pixel P(2,n), being located in a spatial neighborhood of the particular pixel P(1,n), and a fourth value of a fourth pixel P(2,n-1) of the second image of the display sequence of video images, corresponding to the second pixel P(2,n), is less than a second predetermined threshold T2;
- a third evaluation unit 306 for testing whether a third difference D12 between the first value of the particular pixel P(1,n) and the third value of the second pixel P(2,n) is less than a third predetermined threshold T3; and
- a combining unit 308 for establishing that the particular pixel represents the part of the graphics overlay if the first difference D11 is less than the first predetermined threshold T1, the second difference D22 is less than the second predetermined threshold T2 and the third difference D12 is less than the third predetermined threshold T3.

The overlay detection unit 300 optionally comprises a memory device 310 for temporarily storage of video images and intermediate results. Alternatively, the memory device 310 is shared with other components of the apparatus to which the overlay detection unit 300 belongs.

The output of the overlay detection unit 300 is provided at the output connector 314. Preferably, the output is a binary bitmap per input image indicating per pixel of the input image whether that pixel corresponds to a part of a graphics overlay or not. Alternatively, the output is a multi-dimensional signal comprising appropriate color or

luminance values for those pixels for which has been established that they represent a part of a graphics overlay.

The pixels of the images of the display sequence of video images are evaluated in a scanning approach. This scanning might be line by line or column by column. Each pixel for which holds that its value is temporarily stationary and substantially equal to a value of a second pixel of which its value is also temporarily stationary, is labeled as representing a part of a graphics overlay.

Preferably, pixels of consecutive images having mutually equal coordinates are compared. Alternatively, pixels of consecutive images being related by motion vectors are compared. Hence, "corresponding" does not necessarily mean equal coordinates.

The first evaluation unit 302, the second evaluation unit 304, the third evaluation unit 306 and the combining unit 308 may be implemented using one processor. Normally, these functions are performed under control of a software program product. During execution, normally the software program product is loaded into a memory, like a RAM, and executed from there. The program may be loaded from a background memory, like a ROM, hard disk, or magnetically and/or optical storage, or may be loaded via a network like Internet. Optionally an application specific integrated circuit provides the disclosed functionality.

Fig. 3B schematically shows an alternative embodiment of the overlay detection unit 301 according to the invention. The input and output of this embodiment of the overlay detection unit 301 is equal to the input and output of the embodiment of the overlay detection unit 300 as described in connection with Fig. 3A. This embodiment of the overlay detection unit 300 is arranged to detect whether the pixels represent respective parts of graphics overlays on basis of three consecutive images.

The overlay detection unit 301 comprises:

- a first evaluation unit 302 for testing temporal stability, e.g. whether a first difference D11 between a first value of the particular pixel P(1,n) and a second value of a corresponding pixel P(1,n-1) of a second image of the display sequence of video images is less than a first predetermined threshold T1 and whether a fourth difference D11p between the first value of the particular pixel P(1,n) and a first value of a corresponding pixel P(1,n+1) of a third image of the display sequence of video images is less than the first predetermined threshold T1;
- a third evaluation unit 306 for testing homogeneity, e.g. whether a third difference D12 between the first value of the particular pixel P(1,n) and the third value of the

second pixel P(2,n) is less than a third predetermined threshold T3 and whether a fifth difference D13 between the first value of the particular pixel P(1,n) and the fifth value of a third pixel P(3,n) is less than a third predetermined threshold T3; and

- a combining unit 308 for establishing that the particular pixel represents the part of the graphics overlay on basis of the output of the evaluation units 302 and 306.

The overlay detection unit 301 optionally comprises a motion estimation unit 316 for computing motion vectors. Alternatively, the motion estimation unit 316 is shared with other components of the apparatus to which the overlay detection unit 301 belongs. The motion estimation unit 316 is e.g. as specified in the article "True-Motion Estimation with 3-D Recursive Search Block Matching" by G. de Haan et. al. in IEEE Transactions on circuits and systems for video technology, vol. 3, no. 5, October 1993, pages 368-379.

The combining unit 308 is connected to the motion estimation unit 316 to receive information about motion. This information is used by the combining unit 308 as will be explained below. The combining unit 308 is also connected to the memory device 310 for temporarily storage of intermediate results.

The working of the overlay detection unit 301 is as follows. The first step is the stationary test and is performed by the first evaluation unit 302. This step is to initialize the overlay detection unit 301 with potentially static pixels. All pixels which have been static during the three video images and for which the motion estimation unit 316 has found a zero vector are labeled with a value of 1. To be more precise, Equation 2 specifies which pixels $\vec{p}$ are labeled with a value of 1. All other pixels are labeled with a value of 0. The check on being static is based on function $S(\vec{p},\alpha)$ as specified in Equation 1.

$$S(\vec{p},\alpha) = (\max_{i,j=n-1}^{n+1} \left| f(\vec{p},i) - f(\vec{p},j) \right|) < \alpha \qquad (1)$$

The function $S(\vec{p},\alpha)$ checks against a first predetermined threshold α to determine if the luminance value of pixel $\vec{p}$ has remained stable over time, i.e. substantially the same value for three consecutive images. $f(\vec{p},n)$ is the luminance for pixel $\vec{p}$ in image n.

$$I(\vec{p}) = \begin{cases} 1 \quad if \quad \vec{d}(\vec{p}) = \vec{0} \wedge S(\vec{p},\alpha) \\ \quad 0 \quad otherwise \end{cases}$$

with $\vec{d}(\vec{p})$ the motion vector for pixel $\vec{p}$. The zero-vector check is optional, but experiments have proved that the results are significantly less robust without the check. For the zero-vector check, the vector field of image n-1, n or n+1 can be used.

The next step is the shrinking step, based on a homogeneity test and is performed by the third evaluation unit 306. The shrinking step removes the falsely detected pixels. For every pixel $\vec{p}$ labeled with a value of 1, a region of connected pixels $\vec{q}$ is determined for which holds:

- pixel $\vec{q}$ is within a spatial neighborhood of N*M pixels. Typically N*M=21*21 pixels;
- the pixel values of $\vec{p}$ and $\vec{q}$ are substantially mutually equal. Hence, $|f(\vec{p})-f(\vec{q})|<\beta$. This similarity measure can be calculated using luminance, or both luminance and chrominance; and
- $\vec{q}$ is connected to $\vec{p}$ by a path of pixels which adhere to the above two conditions. This condition is not strictly necessary. Without it, the detection method is less robust, but also significantly less computationally expensive.

If for the majority of these pixels $S(\vec{q},\alpha)$ holds and for the rest of these pixels $S(\vec{q},3\alpha)$ holds, the pixel $\vec{p}$ remains labeled with a value of 1. Otherwise, $\vec{p}$ will be labeled with a value of 0.

Preferably, a growing step is performed as a next iteration over the image. The growing step expands the static regions to pixels which are similar to the static regions. Every pixel p which is not labeled with a value of 1 is considered. Similar as specified for the shrinking step a region of connected pixels $\vec{q}$ is determined for which holds:

- pixel $\vec{q}$ is within a spatial neighborhood of N*M pixels. Typically N*M=21*21 pixels; and
- the pixel values of $\vec{p}$ and $\vec{q}$ are substantially mutually equal. Hence, $|f(\vec{p})-f(\vec{q})|<\beta$.

Fig. 4 schematically shows an embodiment of the image processing apparatus 400 according to the invention, comprising:

- a receiving unit 402 for receiving a video signal representing a display sequence of video images;
- an overlay detection unit 408, as described in connection with any of the Figs. 3A and 3B;

- an image processing unit 404 for calculating a sequence of output images on basis of the display sequence of video images and on basis of a graphics overlay detection signal being provided by the overlay detection unit 408; and
- a display device 406 for displaying the sequence of output images as provided by the image processing unit.

The video signal may be a broadcast signal received via an antenna or cable but may also be a signal from a storage device like a VCR (Video Cassette Recorder) or Digital Versatile Disk (DVD). The signal is provided at the input connector 410. The image processing apparatus 400 might e.g. be a TV. Alternatively the image processing apparatus 400 does not comprise the optional display device but provides the output images to an apparatus that does comprise a display device 406. Then the image processing apparatus 400 might be e.g. a set top box, a satellite-tuner, a VCR player, a DVD player or recorder. Optionally the image processing apparatus 400 comprises storage means, like a hard-disk or means for storage on removable media, e.g. optical disks. The image processing apparatus 600 might also be a system being applied by a film-studio or broadcaster.

It should be noted that the above-mentioned embodiments illustrate rather than limit the invention and that those skilled in the art will be able to design alternative embodiments without departing from the scope of the appended claims. In the claims, any reference signs placed between parentheses shall not be constructed as limiting the claim. The word 'comprising' does not exclude the presence of elements or steps not listed in a claim. The word "a" or "an" preceding an element does not exclude the presence of a plurality of such elements. The invention can be implemented by means of hardware comprising several distinct elements and by means of a suitable programmed computer. In the unit claims enumerating several means, several of these means can be embodied by one and the same item of hardware.